Essay

The sum of the parts

Graeme Mitchison

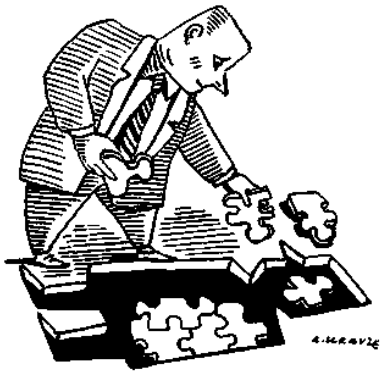
Thirty years ago, Francis Crick, in his brilliant and combative little book *Of Molecules and Men* (Seattle: University of Washington; 1966), expounded what could be regarded as the biological reductionist's credo. The ultimate aim of a modern biologist, he said, should be to explain all biology in terms of physics and chemistry; in particular, it should be possible to explain the development of organisms and the function of higher nervous systems in these terms. The book was combative because he had an enemy to deal with: the organicists and vitalists who were plentiful in those days. It's probably fair to say that the battle against vitalism has been won, at any rate until another dark age engulfs us. But there are other forms of dissension from the outlook he expressed, and one of the aims of a recent CIBA Foundation Symposium — The Limits of Reductionism in Biology (London, 13–15 May 1997) — was to explore these.

Those gathered at the meeting ranged from ardent pro-reductionists — our chairman Lewis Wolpert (University College London) being a distinguished example — to those with a large anti-reductionist axe to grind. In between were persons of less determinate disposition, many of them perhaps not even sure how reductionism should be defined. Fortunately, there was no shortage of definitions; by the end of the meeting, as many had been offered as there were bodies in the room, or perhaps rather more.

William Quinn, a physiologist who studies memory at the Massachusetts Institute of Technology, suggested that anyone who worked with a simplified model system and believed that Nature was simple at heart was a reductionist. What he meant by simplification here was not entirely obvious: "Pavlov reduced dogs," he said, meaning that Pavlov treated his animals as stimulus–response machines, ignoring the fixations they developed towards pieces of experimental apparatus.

From a common-sense standpoint, reductionism means taking things apart in order to understand how they work. The 'understanding' aspect needs to be emphasized, as it is a frequent jibe against reductionism that it consists only in taking apart without any attempt at synthesis. Both components are needed. As Crick put it in *Of Molecules and Men*: "it would be difficult to deduce the detailed functions of a watch either from the intact, unopened running watch, or from the smashed pieces, but a combination of these two approaches would tell us most things about the mechanism."

It is, nevertheless, not entirely clear how much synthesis or integration is permitted before the subject moves into a realm beyond reductionism. Is statistical thermodynamics a reductionist subject? It has its reduced entities, namely particles, but a property like temperature — as the chemist Bob Williams (University of Oxford) repeatedly reminded us — is an ensemble property, not deducible from a single particle. Biological theory at present seems in no danger of losing sight of its constituents in this way, but a certain degree of abstraction is needed to grasp the behaviour of complicated systems. For example, Denis Noble (University of Oxford) has for many years studied the physiology of heart muscle cells in an incontestably reductionist mode, but is currently attempting to integrate the information he has obtained and model a region of the intact heart. His spectacular computer simulations yield insights not attainable by studying the individual components. "Am I a reductionist?" he asked, evidently anticipating a negative



response. "I've not a shadow of doubt you are," replied Wolpert.

Having failed to achieve unanimity on this matter of synthesis, we went on to do the same with levels of organization. No one thinks that it is sensible to try to reduce an organism to atoms; instead, one hopes to study it at several levels, each reducible to the preceding one. Take a frog snapping at a fly: one can analyse the synaptic pathway from retina to muscles, the molecular components of synapses can be teased apart, and the structures of these components can carry us to the level of atoms. But one would not try to put the whole structure together atom by atom.

Are there levels which are in some sense not attainable from those below? This would certainly constitute a limit on reductionism. From time to time, members of our group would deliver jobations about the impossibility of bringing reductionism to bear on the issue of purpose (such as, why the frog snaps at the fly), and these would be received with acquiescence. It's not clear, however, that this type of limitation pertains particularly to reductionism. Purpose carries us into questions of evolution, and evolution is part science and part history.

While these matters were being debated, Robert May (University of Oxford, and UK Government Chief Scientist) dropped in from the corridors of power to tell us about population biology. Muttering "I don't know what I'm doing here," he told us, not without a touch of

impatience, that one should work at whatever level was practical, and reduce as far as was possible. This sound advice leads one to speculate about the future of reductionism. What is practical, what is possible? Will we really be able to put together all the pieces being so energetically produced by the great laboratory-factories of the world?

The scientific world seems to divide into those who think that tasks which are too difficult for the present, like modelling large-scale cellular dynamics or folding proteins, must always be unachievable, and those who see it as only a matter of time before they become routine. History suggests it is safer to side with the latter view: heavier-than-air flying machines seem to be doing quite well nowadays, there are photographs of atoms in the journals every week, and so on. It was good, though, to have some dissent throughout the meeting. Judged by the criterion of liveliness, this was a four-star meeting.

Address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

Correspondence

Widespread eukaryotic sequences, highly similar to bacterial DNA polymerase I, looking for functions

Erik L.L. Sonnhammer and John C. Wootton

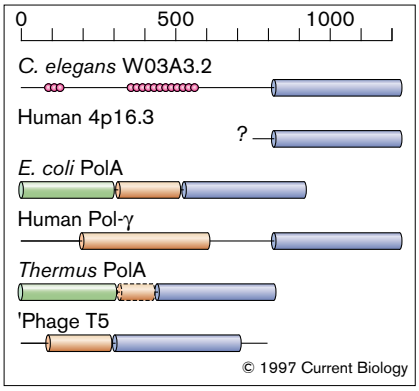
The sequence databases now contain several particularly interesting DNA polymerase sequences for which functions have not yet been assigned. These sequences are present and transcribed in widespread eukaryotes. They are strongly similar in their polymerase domains to the familiar DNA polymerase I (PolA family) of the bacterial kingdom, and are only weakly similar to Pol-γ, the enzyme responsible for mitochondrial DNA replication. Two distinct human sequences are so far available: first, a cDNA of a class that is also represented by *Caenorhabditis elegans* (nematode) genomic and cDNA sequences; second, an intriguing set of rather cryptic short exon sequences interspersed in the 4p16.3 (Huntington's disease) region [1] of the genome. Partial cDNAs, encoding parts of similar DNA polymerase domains, are also available from maize, *Zea mays*, and the malarial parasite *Plasmodium falciparum*. Similar sequences are notably absent, however, from the complete genome sequence of the yeast *Saccharomyces cerevisiae*.

Different classes of DNA-dependent DNA polymerases (EC 2.7.7.7) have well-established roles in DNA replication, repair, mutagenesis and recombination, and share structural similarity in their polymerase domains. The new sequences analysed here clearly belong to family A, as classified by polymerase domain sequence homology [2,3]. Family A polymerases have functions in: DNA repair, and RNA primer removal during lagging strand replication in bacteria (the PolA enzymes); DNA replication in eukaryotic mitochondria (the nuclear-encoded Pol-γ group); and viral DNA replication in several bacteriophages [4,5]. All of these have 3′→5′ and/or 5′→3′ exonuclease domains on the amino-terminal side of the polymerase domain (Fig.1).

The most complete example of the new polymerase sequences is a conceptual translation from the *C. elegans* genome (coding region W03A3.2, Genbank U50184). This contains, in addition to the carboxy-terminal polymerase domain, a long amino-terminal region that does not contain any recognizable sequence similarity to known exonucleases. Six independent cDNA clones are also available from *C. elegans*, covering both the amino-terminal and polymerase regions. No expressed sequence tags (ESTs) from other organisms match the amino-terminal region. The two human genes of this family are distantly related to each other. One is a partial cDNA (dbEST accession W00829), and is more similar in sequence to the *C. elegans* homolog than to the other human gene. The second human gene has been partially assembled conceptually by sequence similarity from 13 short exons interspersed over approximately 130 kilobases of a 2 megabase contig from the Huntington's disease region of 4p16.3.

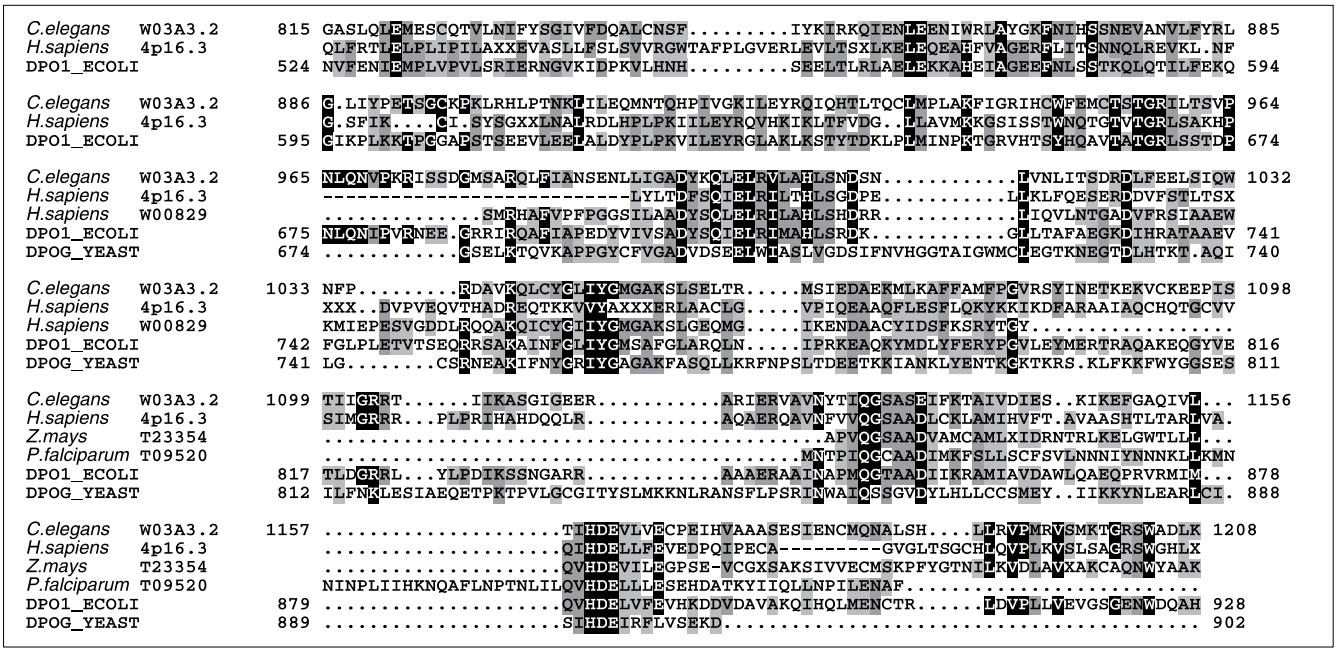
Figure 2 shows the multiple alignment of these sequences and other DNA polymerase family A representatives, including the new *Z. mays* and *P. falciparum* partial cDNAs.

Figure 1



The DNA polymerase domain in family A occurs in combination with various other domains. DNA polymerase domains are colored blue, 3′→5′ exonuclease domains orange, and 5′→3′ exonuclease domains are green. In *Thermus*, an inactive 'remnant' of the 3′→5′ exonuclease domain is present [12] (dotted box). Swissprot accessions: *E. coli* PolA: P00582; Human Pol-γ: P54098; *Thermus aquaticus* PolA: P19821; Bacteriophage T5: P19822. The amino-terminal portion of W03A3.2 lacks clear similarity to other sequences and contains two low-complexity regions (pink circles), which potentially divide it into three globular subdomains.

Figure 2



Alignment of the DNA polymerase domains found in *C. elegans* W03A3.2, the human 4p16.3 gene, *E. coli* DNA polymerase I (Swissprot: DPO1_ECOLI/P00582), *S. cerevisiae* Pol-γ (Swissprot: DPOG_YEAST/P15801) and ESTs from human, maize and *Plasmodium*. The EST names indicate Genbank accession numbers. The two dashed segments in the 4p16.3 gene indicate exons which are presumed necessary, but for which no likely candidates were found based on sequence similarity. X denotes uncertain exon boundaries. The dash in T23354 indicates a frameshift relative to the EST sequence. The multiple alignment was constructed manually based on BLAST2 [13] and TBLASTN [14] alignments. The probability of observing the similarity to PolA by chance, computed using the BLASTP program [14], was $<10^{-50}$ for W03A3.2 and the 4p16.3 gene, and $<10^{-6}$ for each of the EST translations, when searched against the NCBI NR database. The exon structure of W03A3.2 in Genbank was altered according to EST evidence.

All these sequences contain the characteristic conserved residue patterns found in active DNA polymerase domains [6], and show a much greater similarity to PolA than does Pol-γ. Phylogenetic analysis indicates that both *C. elegans* W03A3.2 and the human 4p16.3 sequences are placed among the deeper branches of the PolA and bacteriophage sequences (not shown). Pairwise alignments using these eukaryotic sequences show only very remote homology to the mitochondrial Pol-γ group. The question of the evolutionary origins of these apparently ancient DNA polymerase lineages is completely open.

Recent bacterial contamination or recent transfers from bacterial or bacteriophage sources is evidently ruled out by the presence of these genes in diverse eukaryotes and by their characteristic exon–intron genomic structures. Also, analysis by the Zinfo program [7] showed that the compositional statistics of the W03A3.2 coding sequence are typical for *C. elegans* but atypical for *E. coli*. The phylogeny and distribution, including the occurrence of at least two distantly related human paralogs, could be consistent with multiple ancient parallel transfer events as well as with a single ancient transfer, perhaps through the mitochondrial line, followed by more recent gene duplications. Some lineages, including Pol-γ, may have undergone phases of rapid evolution, together with loss of exonuclease domains and gain of other domains. Presumably, given the wide distribution in protist, plant and metazoan organisms, loss of these polymerase genes occurred in the ancestry of *S. cerevisiae*.

What are the possible functions of these DNA polymerases? Experimental verification of polymerase activity is the next step, but in the interim it is tempting to consider DNA repair. It may be pertinent to investigate the repair of damage to mitochondrial DNA, which is poorly understood but which may have roles in the progression of cancer, diabetes and other chronic diseases [8–10]. A repair function is also consistent with the absence of these sequences in *S. cerevisiae*, as there are well-established differences in repair between this yeast and mammalian cells [11].

References

1. Baxendale S, MacDonald ME, Mott R, Francis F, Lin C, Kirby SF, *et al.*: **A cosmid contig and high resolution restriction map of the 2 megabase region containing the Huntington's disease gene.** *Nat Genet* 1993, **4**:181–186.
2. Ito J, Braithwaite DK: **Compilation and alignment of DNA polymerase sequences.** *Nucleic Acids Res* 1991, **19**:4045–4057.

3. Braithwaite DK, Ito J: **Compilation, alignment, and phylogenetic relationships of DNA polymerases.** *Nucleic Acids Res* 1993, **21:**787–802.
4. Kornberg A, Baker TA: *DNA Replication*, 2nd edn. New York: WH Freeman; 1992.
5. Ye F, Carrodeguas JA, Bogenhagen DF: **The γ subfamily of DNA polymerases: cloning of a developmentally regulated cDNA encoding *Xenopus laevis* mitochondrial DNA polymerase γ.** *Nucleic Acids Res* 1996, **24:**1481–1488.
6. Blanco L, Bernad A, Blasco MA, Salas M: **A general structure for DNA-dependent DNA polymerases.** *Gene* 1991, **100:**27–38.
7. Scherer S, McPeek MS, Speed TP: **Atypical regions in large genomic DNA sequences.** *Proc Natl Acad Sci USA* 1994, **91:**7134–7138.
8. Ames BN: **Endogenous DNA damage as related to cancer and aging.** *Mutat Res* 1989, **214:**41–46.
9. Shen CC, Wertelecki W, Driggers WJ, LeDoux SP, Wilson GL: **Repair of mitochondrial DNA damage induced by bleomycin in human cells.** *Mutat Res* 1995, **337:**19–23.
10. Driggers WJ, Grishko VI, LeDoux SP, Wilson GL: **Defective repair of oxidative damage in the mitochondrial DNA of a xeroderma pigmentosum group A cell line.** *Cancer Res* 1996, **56:**1262–1266.
11. Lindahl T: **Recognition and processing of damaged DNA.** *J Cell Sci* 1995, **19 (Suppl):**73–77.
12. Kim Y, Eom SH, Wang J, Lee DS, Suh SW, Steitz TA: **Crystal structure of *Thermus aquaticus* DNA polymerase.** *Nature* 1995, **376:**612–616.
13. Altschul SF, Gish W: **Local alignment statistics.** *Methods Enzymol* 1996, **266:**460–480.
14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403–410.

Address: Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, Building 38A, National Institutes of Health, Bethesda, Maryland 20894, USA.
E-mail: sonnhammer@ncbi.nlm.nih.gov; wootton@ncbi.nlm.nih.gov
Correspondence: Erik LL Sonnhammer

Barth syndrome may be due to an acyltransferase deficiency

Andrew F. Neuwald

Barth syndrome is an X-linked inherited disorder characterized by short stature, cardioskeletal myopathy, neutropenia, abnormal mitochondria, and respiratory-chain dysfunction [1,2]. It is often fatal in childhood due to cardiac failure or sepsis arising from agranulocytosis. The phenotype associated with this disorder is quite variable, however, and other X-linked cardiomyopathies [3–5] may be allelic to Barth syndrome, which maps to a gene-rich region of Xq28 [6]. Recently, a gene mutated in patients afflicted with Barth syndrome (*G4.5*) was cloned and sequenced [7]. It encodes several proteins (designated tafazzins) by means of alternate splicing. The biological function of tafazzins is unclear; a BLAST search [8] finds significant pairwise similarity only to two hypothetical proteins: one from worm (g1130664), and one from yeast (g1066481).

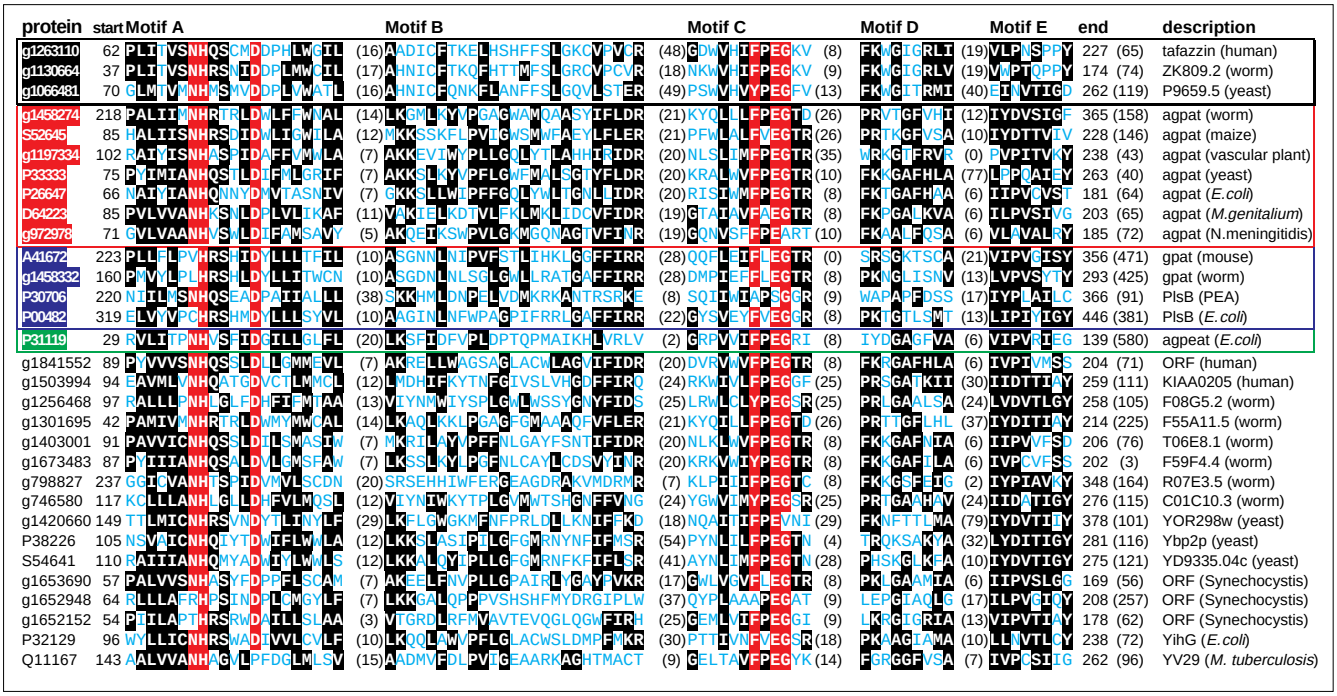
Here, I report that human tafazzins belong to a superfamily consisting of acyltransferases involved in phospholipid biosynthesis and other proteins of unknown function. This superfamily was found using PROBE [9], an automated search and multiple alignment program based on iterative database searches. Starting with a plant 1-acylglycerol-3-phosphate acyltransferase (EC 2.3.1.51), g1197334, PROBE returned a superfamily that includes known and putative acyltransferases from bacteria, fungi, plants, and vertebrate and invertebrate metazoans. Characterized enzymes in this superfamily all function in phospholipid biosynthesis and have either glycerolphosphate, 1-acylglycerolphosphate, or 2-acylglycerolphosphoethanolamine acyltransferase activity.

The sequence alignment contains five conserved regions that presumably reflect similar structural and functional features shared by these proteins (Fig. 1). As all of the characterized proteins are acyltransferases involved in phospholipid biosynthesis, the uncharacterized proteins are likely to have similar catalytic activity. Notably, motif A contains a fully conserved residue position that may correspond to a catalytic histidine, as has been found at the active site of other CoA-dependent hydrolases [10]. Of course, it is possible that tafazzins perform some other hydrolytic function. Indeed, hydrolytic activity was previously predicted for tafazzins based on weak similarity to the *Escherichia coli radC* gene [11], which may possess hydrolytic activity needed for DNA repair. Nevertheless, the more extensive similarity of tafazzins to these acyltransferases implies a closer similarity in function.

The potential acyltransferase activity of tafazzins suggests a possible disease mechanism underlying Barth syndrome. Differential splicing of tafazzins [7] and the existence of at least nine of these putative acyltransferases in roundworm — by contrast with the four detected in the *E. coli* genome — suggests that a variety of substrate-specific or tissue- and organelle-specific forms of these acyltransferases exist in eukaryotes. If so, then the mitochondrial structural and respiratory-chain abnormalities associated with Barth syndrome may be due to alterations in mitochondrial membrane phospholipid composition. Consistent with this notion, a temperature-sensitive Chinese hamster ovary cell mutant deficient in an enzyme needed for cardiolipin biosynthesis showed alterations in mitochondrial morphology and respiration [12].

It is important to note that the roundworm ZK809.2 gene may be an ortholog of the human *G4.5* (tafazzins) gene. It shares several splice sites with the human gene and the predicted product is more closely related to tafazzins than to any other protein in the superfamily. Notably, the worm protein is missing exon 5, which appears to be removed from many of the tafazzin splice variants [7]. Furthermore, the worm homolog shares several conserved regions with human tafazzins that are

Figure 1



Alignment of representative sequences in the tafazzins, or acyltransferase superfamily. A total of 53 proteins in the NCBI non-redundant database were detected by the PROBE search [9], which used default parameter settings. Conserved residues are highlighted in red (for the most conserved positions) or black. Numbers in parentheses are gap lengths. Human tafazzin is detected at the $p < 0.00001$ level of significance; this is based on a database search using an alignment lacking sequences with statistically significant pairwise similarity to tafazzins (that is, lacking sequences g1263110, g1130664, g1066481, g1841552, g1403001 and g1673483). (The database search and the *p*-value calculation were done as previously described [13].) Protein identifiers are highlighted according to the following color scheme: black, tafazzin and close homologs; red, 1-acylglycerol-phosphate acyltransferases (agpat); blue, glycerolphosphate acyltransferases (gpat); green, 2-acylglycerolphosphoethanolamine acyltransferase (agpeat); unhighlighted, proteins of unknown function.

unconserved in the superfamily as a whole. Thus, ZK809.2 mutants may serve as a useful model to explore the molecular mechanisms underlying Barth syndrome.

References

1. Barth PG, Van den Bogert C, Bolhuis PA, Scholte HR, van Gennip AH, Schutgens RB, Ketel AG: **X-linked cardioskeletal myopathy and neutropenia (Barth syndrome): respiratory-chain abnormalities in cultured fibroblasts.** *J Inherit Metab Dis* 1996, **19:**157–160.
2. Barth PG, Scholte HR, Berden JA, Van der Klei-Van Moorsel JM, Luyt-Houwen IE, Van 't Veer-Korthof ET, *et al.*: **An X-linked mitochondrial disease affecting cardiac muscle, skeletal muscle and neutrophil leucocytes.** *J Neurol Sci* 1983, **62:**327–355.
3. Hodgson S, Child A, Dyson M: **Endocardial fibroelastosis: possible X-linked inheritance.** *J Med Genet* 1987, **24:**210–214.
4. Orstavik KH, Skjorten F, Hellebostad M, Haga P, Langslet A: **Possible X-linked congenital mitochondrial cardiomyopathy in three families.** *J Med Genet* 1993, **30:**269–272.
5. Gedeon AK, Wilson MJ, Colley AC, Sillence DO, Mulley JC: **X-linked fatal infantile cardiomyopathy maps to Xq28 and is possibly allelic to Barth syndrome.** *J Med Genet* 1995, **32:**383–388.
6. Bolhuis PA, Hensels GW, Hulsebos TJ, Baas F, Barth PG: **Mapping of the locus for X-linked cardioskeletal myopathy with neutropenia and abnormal mitochondria (Barth syndrome) to Xq28.** *Am J Hum Genet* 1991, **48:**481–485.
7. Bione SP, D'Adamo P, Maestrini E, Gedeon AK, Bolhuis PA, Toniolo D: **A novel X-linked gene, G4.5, is responsible for Barth syndrome.** *Nat Genet* 1996, **12:**385–389.
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403–410.
9. Neuwald AF, Liu JS, Lipman DJ, Lawrence CE: **Extracting protein alignment models from the sequence database.** *Nucleic Acids Res* 1997, **25:**1665–1677.
10. Leslie AG, Moody PC, Shaw WV: **Structure of chloramphenicol acetyltransferase at 1.75-Å resolution.** *Proc Natl Acad Sci USA* 1988, **85:**4133–4137.
11. Mushegian AR, Bassett DE, Boguski MS, Bork P, Koonin EV: **Positionally cloned human disease genes: patterns of evolutionary conservation and new functional motifs.** *Proc Natl Acad Sci USA* 1997, **94:**5831–5836.
12. Ohtsuka T, Nishijima M, Suzuki K, Akamatsu Y: **Mitochondrial dysfunction of a cultured Chinese hamster ovary cell mutant deficient in cardiolipin.** *J Biol Chem* 1993, **268:**22914–22919.
13. Neuwald AF, Liu JS, Lawrence CE: **Gibbs motif sampling: detection of bacterial outer membrane protein repeats.** *Protein Sci* 1995, **4:**1618–1632.

Address: National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland 20894, USA.